Formative Evaluation of a Frame-Based Model of Locative Relationships in Human Anatomy

Carol A. Bean, Ph.D.
Department of Medical Informatics
Columbia University College of Physicians and Surgeons

The verb structure of narrative text in a gross anatomy textbook was analyzed to identify locative relationships. The 169 locative indicators were organized semantically to construct a frame-based model. The validity and coverage of the model was assessed and compared with the UMLS Semantic Net Relations using a novel test set of 71 indicators. All mapped directly to the frame model, while 60% mapped directly to UMLS.

Our conceptual world is most commonly and effectively modeled by representing its component concepts and relationships. However, relatively little attention has been given to the development of a rich set of well-defined relationships for the organization of medical knowledge, especially when compared to the rich set of well-defined concepts that exists.

Locative relationships are basic to human experience and cognition. In the present context, locative relationships are simply specified relationships between entities that locate them in space, one relative to the other. These include both spatial, and by extension, physical relationships. A thorough understanding of locative relationships among anatomic entities is central to the acquisition of medical knowledge, and to the subsequent practice of medicine. Although the Semantic Net of the National Library of Medicine's Unified Medical Language System (UMLS) includes a set of anatomic semantic types, the spatial relationships allowed among them are sparse and may not be adequate to precisely express the exact nature of locative relationships between and among specific anatomic entities (see UMLS documentation[1]). This notion is supported by recent work suggesting the need to add two new relationships for anatomy to the UMLS.[2]

A model of locative relationships among anatomic entities is being developed based on the argument structure of verbs taken from relevant narrative text and clustered according to sense, forming groups of verbs that express the same basic semantic relationship. Such verbs and their associated parts of speech may then be seen to occupy the same semantic field. For example, the verbs in "bronchial arteries *supply* blood to the lungs" and "vagus nerves *carry* respiratory reflex afferents" convey the same spatial sense and thus fall into the same semantic cluster. The relational structures for the semantic groupings are represented as frames. Briefly, the rationale for this approach is as follows. Verbs designate relationships in two ways: the verb sense determines the basic nature or type of relationship and its argument structure determines the roles played by the participants in the relationship. Since roles are tied to the nature of the relationship, verbs that designate the same basic relationship share the same set of roles, or in other words, the same relational structure.[3]

METHODS

Narrative text was analyzed from the textbook of anatomy[4] currently used in the first-year Clinical Anatomy course at Columbia University's College of Physicians and Surgeons. Two chapters were examined: Chapter 12, Pleural Cavities and Lungs, and Chapter 11, Thoracic Cage. Each locative indicator was identified manually by examining verbs and their arguments, specifically verbs, verb phrases, adverbs, and prepositional phrases.

The list of locative indicators obtained was organized into semantic clusters based on grouping verbs judged by the author to have similar or equivalent senses. The verb clusters were assigned class names corresponding to the most general relationship type descriptive of the entire group, and a frame structure was generated for that class, with slots to account for all component participants in the relationship. These class clusters were repeatedly subdivided and reclassified according to the predominant features of the relationship, as best represented by the frames or slots as appropriate. The process ended when no further meaningful divisions could be determined, or when the number of indicators became too small.

The ability of the model to cover a novel test set of locative indicators was compared with that of UMLS. The test set of locative indicators was derived from a third chapter of the anatomy textbook, chosen to represent both a different region and functional system: Chapter 3, Musculoskeletal System. The same procedure described above was used to identify locative indicators in the text and determine their semantic senses. They were then classified and mapped directly to the most specific level of component possible in the relationship schema devised above. The frequency of indicators that mapped to each relationship class or subclass was tabulated. In a similar procedure, the full test set of 71 locative indicators from the test chapter was mapped to the UMLS Semantic Net physical and spatial relationships, and the results compared with those obtained for the frame-based model.

1091—8280/97/$5.00

RESULTS

Semantic clusters and model structure
One hundred specific locative indicators were found in the first chapter examined during the discovery process. The 106 indicators found in the second chapter added another 69 new instantiations to the total, for 169 altogether. The common pattern observed may be generalized as one in which the verb denoted the primary relationship, optionally modified or specified by a prepositional phrase, typically adverbial. In some indicators, the same verb was modified by such closely related adverbs (or phrases) that they were considered as a single case for purposes of the analysis (e.g., *runs by, runs past*); thus totals in the tables will be slightly lower. The frequency and distribution of indicator mappings into frame structures are presented in Table 1. Figure 1 shows the frame hierarchy and selected slot structures.

The conceptually simplest structures were those that indicated the site or relative location of an object, and were described using a **LOCATION** frame, which is a metaphorical specialized **LINK**. Both **LOCATION** and **LINK** describe two entities (**A** and **B**) and the nature of some specified type of relationship between them. In **LINK**, **A** is **Link**ed to **B**, as in "*each lung* **[A]** *is suspended by* **[Link]** *a mesenteric pulmonary ligament* **[B]**." In **LOCATION**, the **Location** of **A** relative to some **B** is given, and optionally indicates the nature of the **Direction** and/or **Distance** between them; for example, "*the neurovascular bundle* **[A]** *lies* **[Location]** *immediately* **[Distance]** *behind the inferior edge* **[Direction]** *of each rib* **[B]**." Altogether, 35 indicators were classified as **LOCATION**s. These were further specified according to the **Direction** and **Distance** characterizing the "gap" between the two entities, or by the relative strength of association or link between them. Three indicators described **LINK**s.

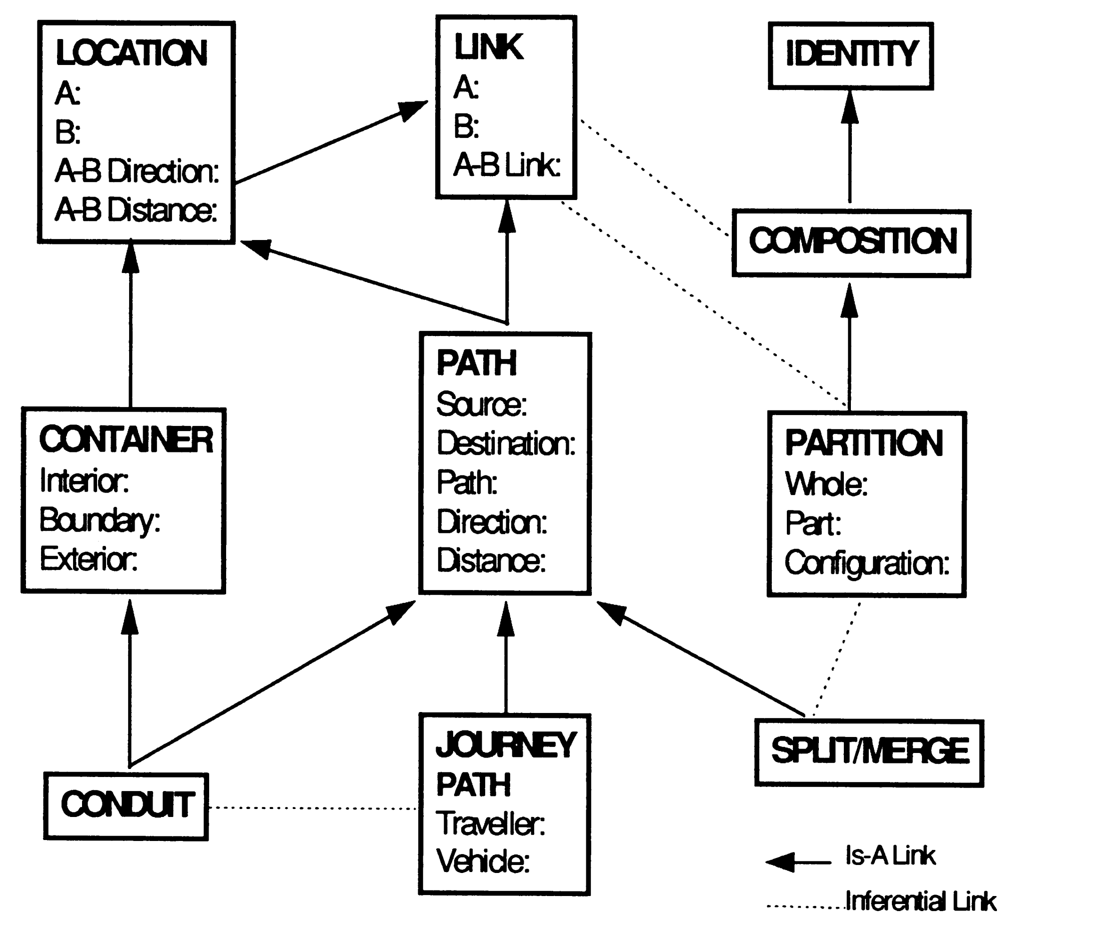

Figure 1. Model of locative relationships showing frame slot structure

Table 1. Frequency of instantiated indicators mapped to each frame or slot type for the relationship discovery ("training") sets and the test set

FRAMES and Slots		Dis1	Dis2	Test	Cum.Tot
IDENTITY		4	11	7	12
	EQUIVALENCE	2	4	2	5
	COMPOSITION	1	2	2	2
	PARTITION	1	5	3	5
LINK		1	3	2	3
LOCATION		15	16	23	35
	Basic	4	4	10	10
	"associated"	3	2	1	5
	"juncture"	2	4	3	6
	"attached"	2	3	3	6
	Distance	1	1	2	2
	Direction	3	2	4	8
PATH		36	30	19	62
	Source	2	1	2	5
	Destination	3	5	1	7
	Path	12	6	5	14
	Distance	2	2	1	3
	Direction	5	5	5	12
	MERGE	4	5	1	8
	SPLIT	6	3	1	6
	(CONDUIT)				
	JOURNEY	2	1	1	3
	Portal	0	2	2	4
CONTAINER		4	12	6	15
	Exterior	1	1	1	2
	Boundary	1	7	1	7
	Interior/Contents	3	4	4	6

Note: Totals do not map to raw scores reported in text, because some instantiations were collapsed when assigned to subclasses

The model has relational structure in that the relationships themselves are interrelated on a higher ("meta") level. For example, some relationships are specializations of others. Frames can be made more specific by placing constraints on slot identities or by adding slots. Composite frames may be formed when one or more simpler frames are combined into a more complex relational structure.

LOCATION +	LINK ->	PATH
A []	A []	Source []
B []	B []	Destination []
----	Link []	Path []
Direction []	----	Direction []
Distance []	----	Distance []

The **PATH** frame is a composite of the more basic **LOCATION** and **LINK** frames. In **PATH**, the **Source** slot can be seen to derive from the **LOCATION-A** and **LINK-A** slots and the **Destination** slot from the **LOCATION-B** and **LINK-B** slots in **LOCATION** and **LINK**, respectively. The **Path** slot is from **LINK**'s **Link** slot, while **Direction** and **Distance** come from **LOCATION**'s **Direction** and **Distance** slots, respectively. Thus, as **PATH** is a specialization of both the **LOCATION** and **LINK** frames, each slot in **PATH** is itself a refinement of the corresponding slots in **LOCATION** and **LINK** from which they derived. For example, "*The pulmonary trunk* **[Path]** *rises* **[Direction]** *from the conus arterius of the right ventricle of the heart* **[Source]** *for approximately 5 cm* **[Distance]**, *where it bifurcates* **[Destination]** *into the left and right pulmonary arteries*." Because this example has two **Destinations**, it also illustrates a specialization of **PATH, SPLIT**. Similarly, a **CONDUIT** may be seen as a **PATH** that is a **CONTAINER**; when the **Contents** move in the **CONDUIT**, the **PATH** becomes a **JOURNEY**, and the **Contents** become the **Traveler.** Certain additional relationships may be inferred without being specified, such as this one, owing to the rich system of inter-relationships among the frame (see Figure). The **JOURNEY** frame incorporates slots to represent a **Traveler**, and the possibility of a **Vehicle**, into the basic structure of the **PATH** frame, and is thus more highly specified. By constraining the number (i.e., greater than one) of **PATH**'s **Destination** and **Source** slots, the **MERGE** and **SPLIT** frames are further specializations of **PATH**, respectively.

The largest group of indicators (62 overall) clustered among the various slots of a **PATH** frame (**Source**, **Destination**, **Path**, **Direction**, and **Distance**) and its specializations (**MERGE/SPLIT**, **CONDUIT**, and **JOURNEY**). Examples of these **PATH** indicators and their slot mappings include: *leaves* and *moves away from* to **Source**, *reaches* and *projects into* to **Destination**, *continues* and *passes along* to **Path**, *drops* and *arches* to **Direction**, and *stretches beyond* to **Distance**. The remaining indicators divided themselves among frames of **IDENTITY** (12 overall) and **CONTAINER** (15 overall).

In addition to the standard anatomic descriptors of orientation and position (e.g., *anterior/posterior*, *proximal/distal*) were found common prepositions such as *above*, *toward*, and *between*. In their simplest locative use, they modify the location of one entity relative to another, as in "synovial cavity *between* the synovial membrane and bone." They are also used in the more complex formulations to specify or focus on one or more conceptual element in the frame, corresponding to a given frame slot or slot value. For example, *extend* is a projection specialization of **PATH.** Adding the prepositional modifiers to form *extends from* and *extends to* emphasizes the **Source** and **Destination** slots, respectively.

Tests of new set of locative indicators

The test chapter contained 71 locative indicators; 33 were already accounted for, having been found in the Discovery chapters. This added another 38 to the previous total of 169, bringing the grand total to 207. All 38 mapped directly to previously identified frame structures in the model; no additional frames or slots were necessary to accommodate the novel instantiations. For example, the novel indicator in "ligaments are *embedded in* bone" was mapped directly to the "**attached**" specialization of **LOCATION**.

The 71 locative indicators from the Test chapter collapsed into 57 distinct ones after equivalent verbs were combined as described above. These were mapped to their equivalent or most closely related UMLS Semantic Net relationships. While it was possible to map all 57 indicators to some UMLS relationship, only somewhat more than half (34) were direct matches; the remaining 23 mappings formed approximate matches with a wide variation in degree of semantic closeness within a relationship category.

Indicators of **COMPOSITION** (n=3), **PARTITION** (n=2), and **LINK** (n=2) matched the equivalent UMLS relationships "composed_of," "part_of," and "interconnect," respectively. UMLS does not contain a general **PATH** relationship, but does have a single relationship , "traverses," that indicates the **Path** or course-traveled component of the **PATH** schema. Thus, only those seven indicators of **PATH.Path** directly matched a UMLS relationship. The remaining twelve **PATH** indicators had to be mapped laterally to this sibling slot type. For **CONTAINER** indicators, **Boundary** (n=1) and **Interior/Contents** (n=4) matched UMLS relationships "surrounds" and "contains," respectively; the single **Exterior** indicator was judged most closely related to "contains." The ten **Basic LOCATION** indicators matched "location_of," as "**juncture**" matched "adjacent_to" (n=3), and "**attached**" matched "connected_to" (n=3). The remaining seven **LOCATION** indicators ("**associative**," **Distance**, **Direction**) and two locative **EQUIVALENCE** indicators seemed to map most closely to UMLS's "location_of."

DISCUSSION

The frame-based model of locative relationships seems both to account for the variety of pertinent relational structures among anatomic entities and to represent them at the desired level of detail and complexity. The structure also permits the expression of modifiers, supports addition of new relationships, and in this assessment was able to accommodate new class structures as necessary.

It should be noted that the generalizability, as well as reliability, of this assessment may be limited by the subjective nature of the approaches used. The frequencies of specific instantiations and their relative distributions among the classes (frames) and subclasses (slots) should also be interpreted with caution, as the limitations of the sample set would almost certainly bias the exact numbers. The organization of the textbook by body region and system limited the analysis to certain types of anatomic entities. This factor may account for some surprises among the results, such as how rare were "part_of" relationships.

The decreasing trend of new relationships with each successive chapter analyzed (100, 69, 38) suggests that the totality of novel relationships has not yet been discovered, but also suggests that it will diminish with each additional, succeeding pass. Even if novel relationships continue to be uncovered, the frame-based model structure should make it a manageable level of effort to continue to incorporate them. It remains to be seen how much additional data must be gathered to discover the full set of locative relational schemas. Certainly the full spectrum of relationships covering the entire domain must be sampled and represented in the model before the breadth of coverage may be assessed with confidence. The exact nature of the relationships may vary somewhat with participation by different types of anatomic structures. However, it seems reasonable to assume that most, if not all, of the pertinent classes may be represented in this version, and that further data will contribute primarily to model refinement and specialization.

UMLS Semantic Net relationships were found to cover the locative relational domain at a generalized level, but because much detail was lost, this system was not as expressive of the complex structure of these relationships as was the model. Each UMLS relationship corresponds to some structural element in the frame model; however, most are at the slot level and not all frame slots are represented by UMLS relationships. Thus, there is more detail in the frame model, and it enables a more complete representation. This suggests the potential value for both a greater number and variety of locative relationships, as well as an alternate arrangement using a different structure. Because physical relationships imply localization, these may better form a subset or specialization of spatial relationships than a sibling class. The structure of the UMLS Semantic Net would permit its extension for additional locative relationships, but not the incorporation of the complex relational structure used in this model.

Perhaps the greatest advantage of using a frame structure for representation of anatomic and other knowledge is that frames by their very nature represent n-ary relationships. The Semantic Net structure of the UMLS allows only binary relationships; more complex relationships must be formed by chaining. Many relationships in the real world, including some in this analysis, are n-ary, and involve more than two participants. For example, the locative relationship pair "separates_from" and "separated_by" implies and requires a minimum of three participant entities: the two entities being kept separate and a third that serves to separate them. Expressing this as a series of three binary relationships not only is difficult, but it fails to express the concurrent interaction among the entities. The frame structure also permits the specialization and combination of parts of different frames or entire frames to create new relational representations.

The absence of a rich formal model of locative relationships hinders our ability to represent spatial knowledge, and to integrate symbolic and geometric forms of representation. This approach could provide the basic framework to develop such a model. Further research is needed to address representation of additional domains to "flesh out" the model, as well as incorporation of logic into the relational structure.

Acknowledgments

Thanks are due Jim Cimino for his helpful review of this paper. This work has also benefitted from discussions with Jim, along with Rebecca Green and Pat Molholt, and was supported by NLM Training Grant LM07079.

References

1. National Library of Medicine. UMLS Knowledge Sources Documentation, 8th Edition. Washington DC: US Department of Health and Human Services, National Institutes of Health, January 1997.

2. Rosse C, BenSaid M, Eno KR, Brinkley JF. Enhancements of anatomical information in UMLS Knowledge Sources. October 1995, Proceedings of the 19th Annual Symposium on Computer Applications in Medical Care, New Orleans LA, 873-7.

3. Green R. A relational thesaurus: Modeling semantic relationships using frames. Final Report to OCLC Online Computer Library Center, January 1997.

4. April EW. Anatomy (2nd edition), National Medical Series, Baltimore:Williams & Wilkins, 1990.